The 20th Century Reanalysis Version 3 Ensemble Data Assimilation System

Jeff Whitaker

NOAA Earth System Research Lab, Boulder, CO, USA
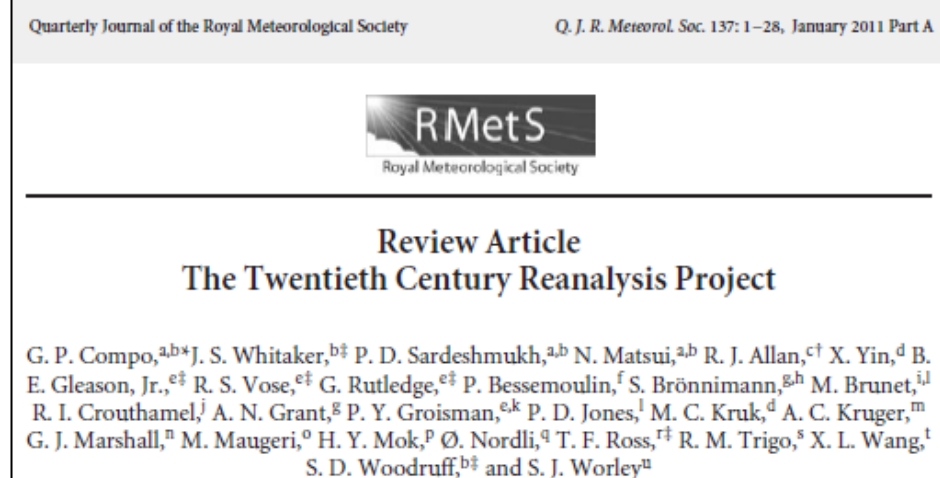
jeffrey.s.whitaker@noaa.gov

Gilbert Compo

CIRES, University of Colorado, Boulder, CO, USA

The 20CR project
(dx.doi.org/10.1002/qj.776)

G. P. Compo,[a,b,*] J. S. Whitaker,[b‡] P. D. Sardeshmukh,[a,b] N. Matsui,[a,b] R. J. Allan,[c†] X. Yin,[d] B. E. Gleason, Jr.,[e‡] R. S. Vose,[e‡] G. Rutledge,[e‡] P. Bessemoulin,[f] S. Brönnimann,[g,h] M. Brunet,[i,l] R. I. Crouthamel,[j] A. N. Grant,[g] P. Y. Groisman,[e,k] P. D. Jones,[l] M. C. Kruk,[d] A. C. Kruger,[m] G. J. Marshall,[n] M. Maugeri,[o] H. Y. Mok,[p] Ø. Nordli,[q] T. F. Ross,[r‡] R. M. Trigo,[s] X. L. Wang,[t] S. D. Woodruff,[b‡] and S. J. Worley[u]

- **Goal**: Use data assimilation to produce the longest possible ***instrument-based*** estimates of global weather and climate for comparison to paleoclimate reconstructions and climate model simulations.

- Uses only surface pressure observations (from the *International Surface Pressure Databank)* in an EnKF assimilation system.

- ECMWF has just completed a similar effort using 24-h window 4DVar (ERA-20C: *http://www.ecmwf.int/en/research/climate-reanalysis/era-20c*)

20CRv2 DA system: 1871-present

- 56 member EnKF using T62L28 version of NCEP GFS.
  - Serial EnSRF, analysis performed in joint observation-model space.
- Localization – 4000km in horizontal, 4 scale heights in vertical.
- Inflation – fixed values for NH,SH tropics that change 3 times (1890,1920 and 1950).
- Observation bias correction – based on last 60 days of O-F.
- QC – background check based on ensemble spread plus buddy check.
- Obs thinned using F-test (If $\mathbf{HP^aHT/HP^bHT}$ is not less than 1 at the 55% level based on the F–test with 56 degrees of freedom, then the observation was not assimilated). 98% retained in 1891, 32% in 2005.

Local Anomaly Correlation of 300 hPa geopotential height anomalies from 20CRv2 vs ERA40 (1979 to 2001)

Correlation 20CR vs ERA40 (300 hPa Height 1979-2001)

Black curves show where NCEP-NCAR and ERA40 correlate > 0.975
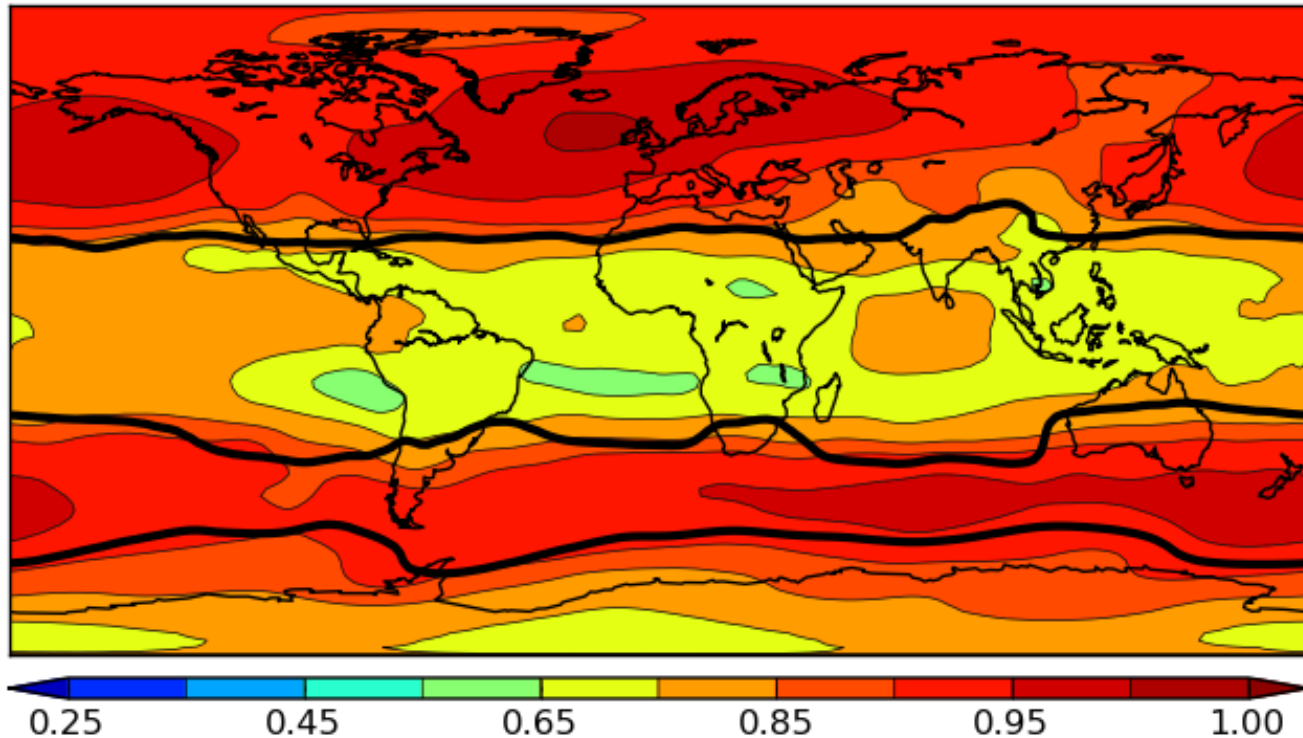
Northern and Southern Hemisphere agreement is when ERA40 has satellite observations. Tropics, not so much.

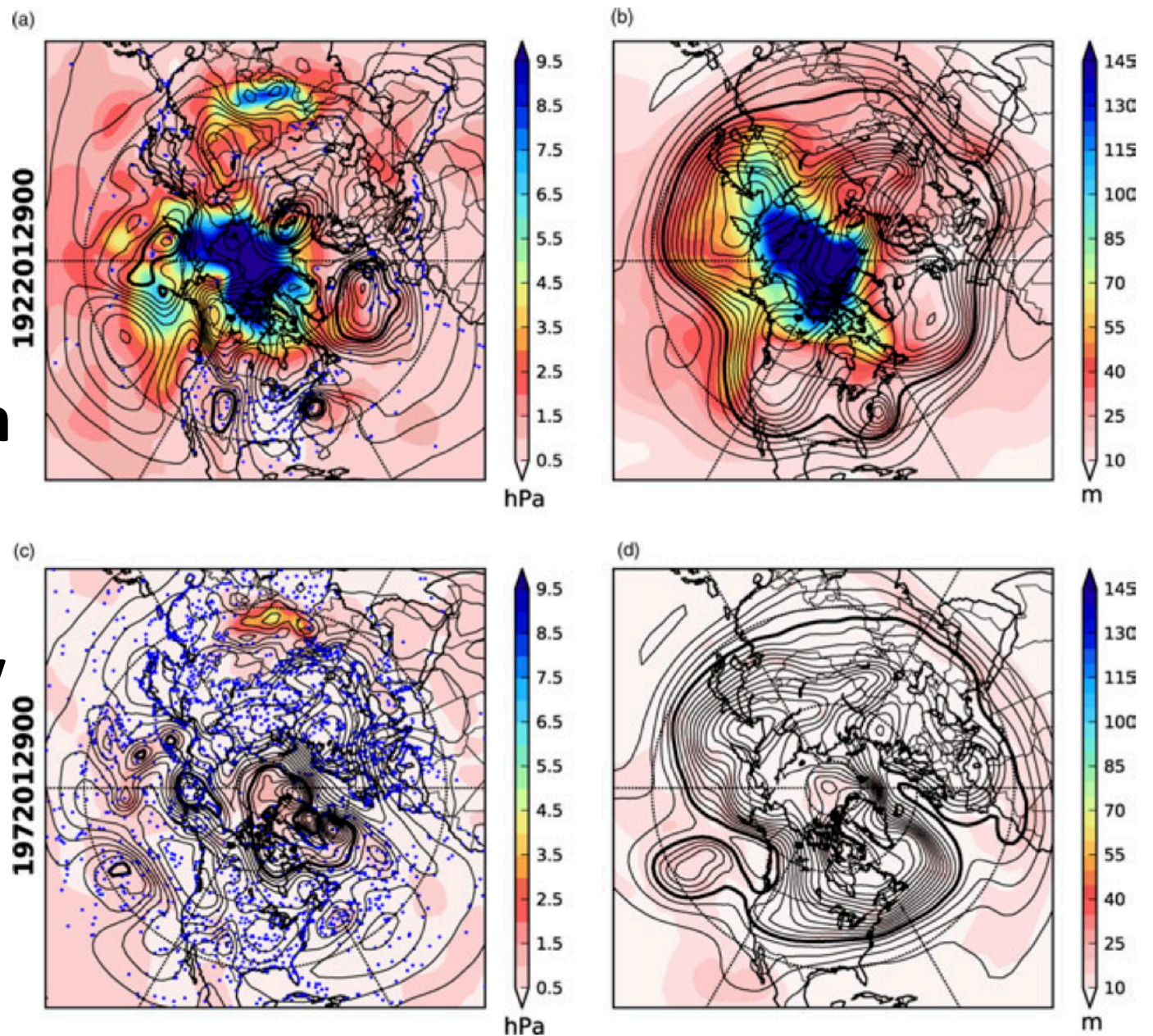**Flow and observation network dependent uncertainty estimate**

Figure 1. Synoptic charts of the ensemble mean analysis and ensemble uncertainty for (a) sea-level pressure and (b) 500 hPa geopotential mean and spread at 0000 UTC on 29 January 1922. (c, d) are as (a, b), but for 29 January 1972. Line contours indicate the analysis and shading indicates the uncertainty measured as the ensemble standard deviation (or spread) at each location. The line contour interval in (a,c) is 4 hPa, with the 1000 hPa contour thickened, and in (b,d) it is 50 m, with the 5600 m contour thickened. The shading interval in (a,c) is 0.25 hPa and in (b,d) is 5 m.

Problems in 20CRv2

- Bug in specification of sea ice causing boundary layer temps too warm in arctic – *fixed in 20CRv2c*.
- Fixed inflation caused too little (too much) spread in data rich (data poor) regions – *new inflation algorithm in version 3.*
- Fixed localization for entire period. Localization probably suboptimal – should depend on observing network – *varying localization algorithm in version 3.*
- Low resolution model (T62L28) – *increased to T254L64 in version 3*. Ens size also increased from 56 to 80.
- Buddy check is yes/no – should give obs with a high probability of having a gross error some weight – *new QC algorithm in version 3 (similar to VarQC) that allows for non-gaussian ob error PDF.*

New covariance inflation

Surface pressure inflation (annual average for 2000)
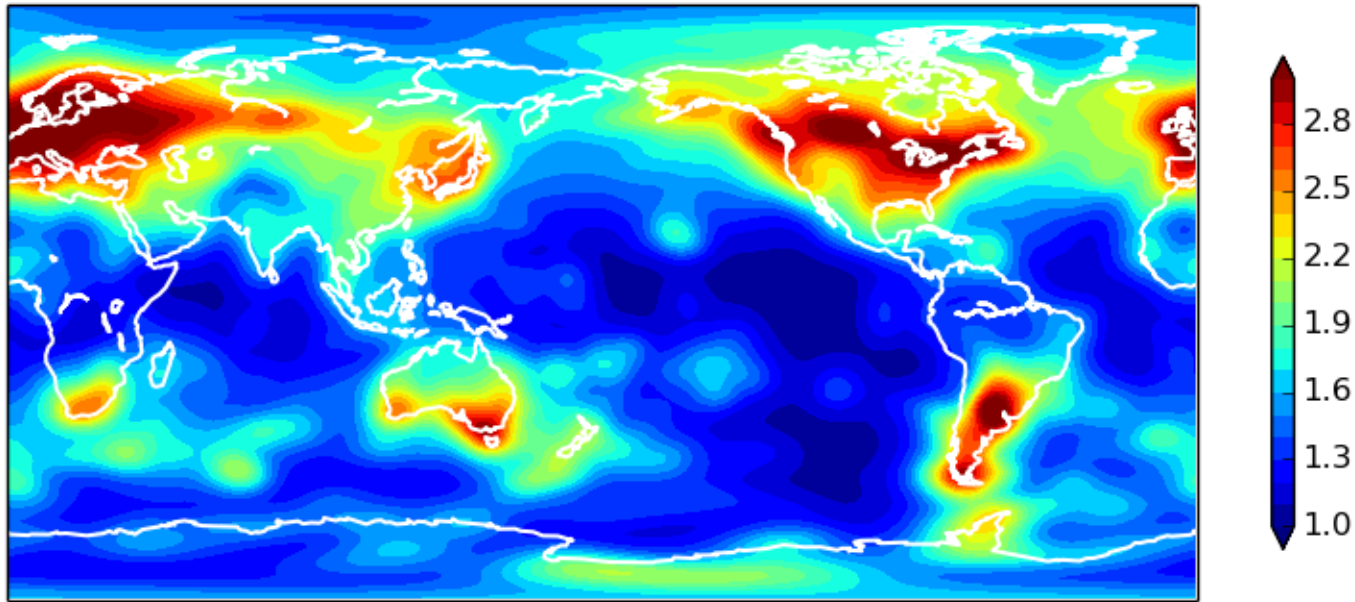
Relaxation to prior spread (RTPS): $\sigma^a \leftarrow \alpha\sigma^b + (1-\alpha)\sigma^a$

which implies $x'^a \leftarrow x'^a [1+\alpha(\sigma^b-\sigma^a)/\sigma^a]$

Here we use $\alpha = 0.9$

Previously, inflation was piecewise constant (NH,TR,SH)

Ref: *journals.ametsoc.org/doi/pdf/10.1175/MWR-D-11-00276.1*

Dealing with non-gaussian ob errors

- Analagous to 'VarQC' in variational systems (Huber norm ob error PDF – gaussian with exponential tails).
- Alternative to 'buddy-check' used in 20CRv2.
- Algorithm based on Dharssi et al 1992 (DOI: 10.1002/qj.49711850709)
  - Iteration solution of serial EnSRF.
  - Ob error modified at each iteration to account for heavy tails.
    - At each iteration, ob error variance multiplied by inverse probability that observation **does not** have a gross error
    - prob. of gross error is linearly proportional to distance from background (from previous iteration).
  - If other surrounding obs support an outlier, probability of gross error will be decreased within the iteration.
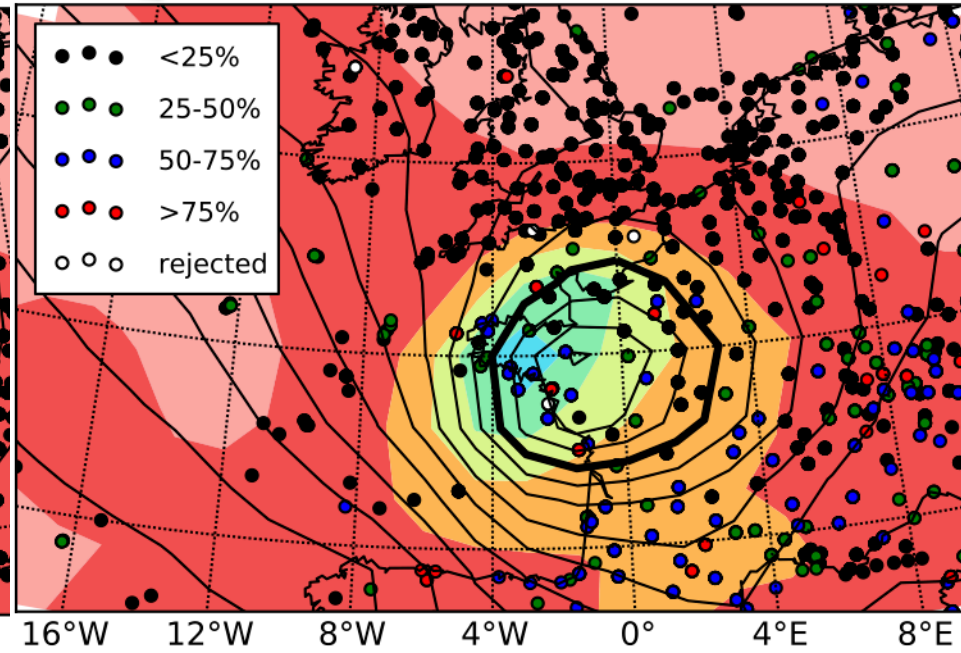
New QC – 'Lothar' storm December 1999

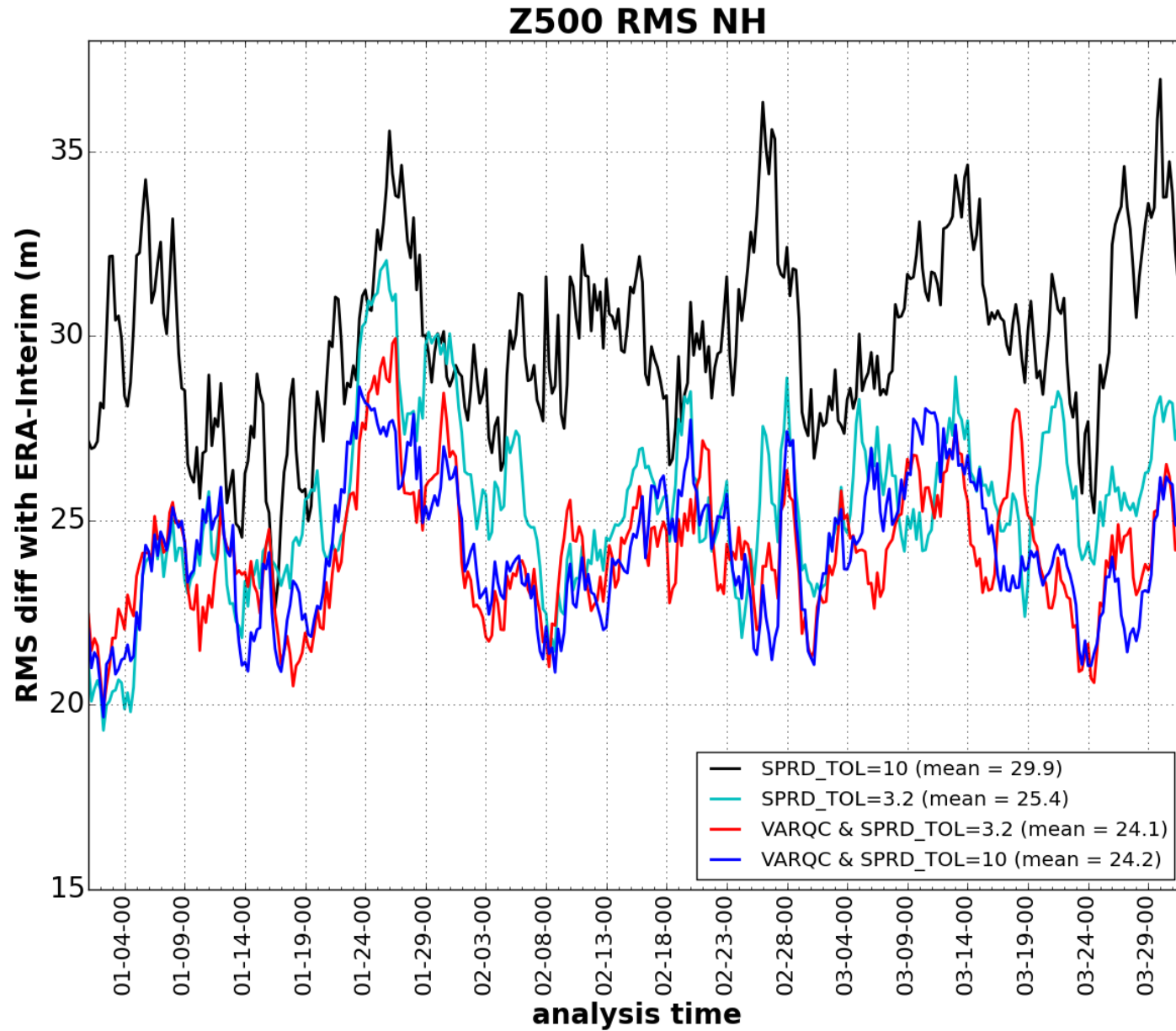probability of gross error

Background check QC only

Huber norm QC

	used
• • •	used
○ ○ ○	rejected

• • •	<25%
• • •	25-50%
• • •	50-75%
• • •	>75%
○ ○ ○	rejected
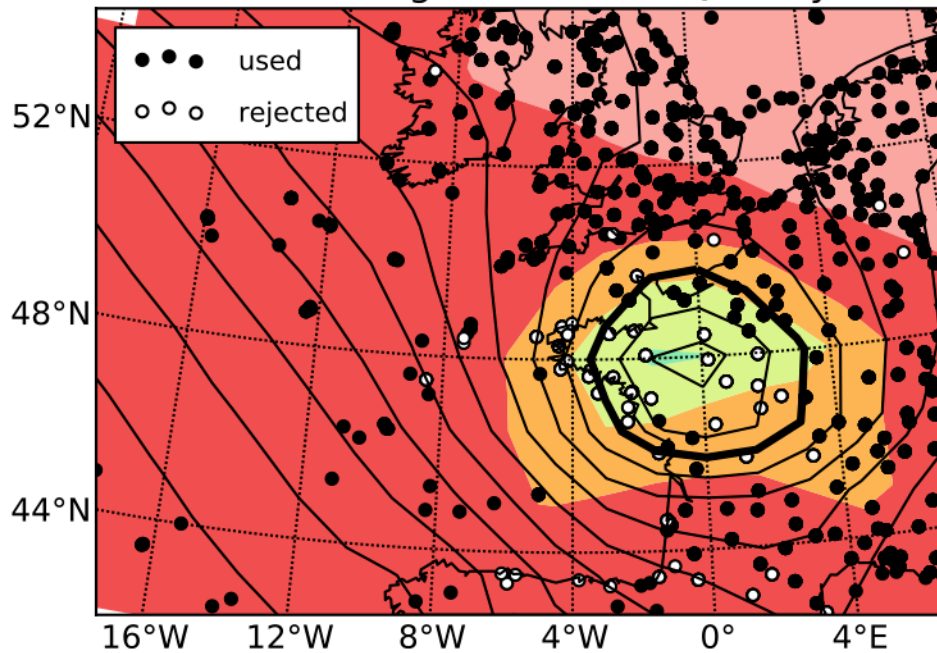
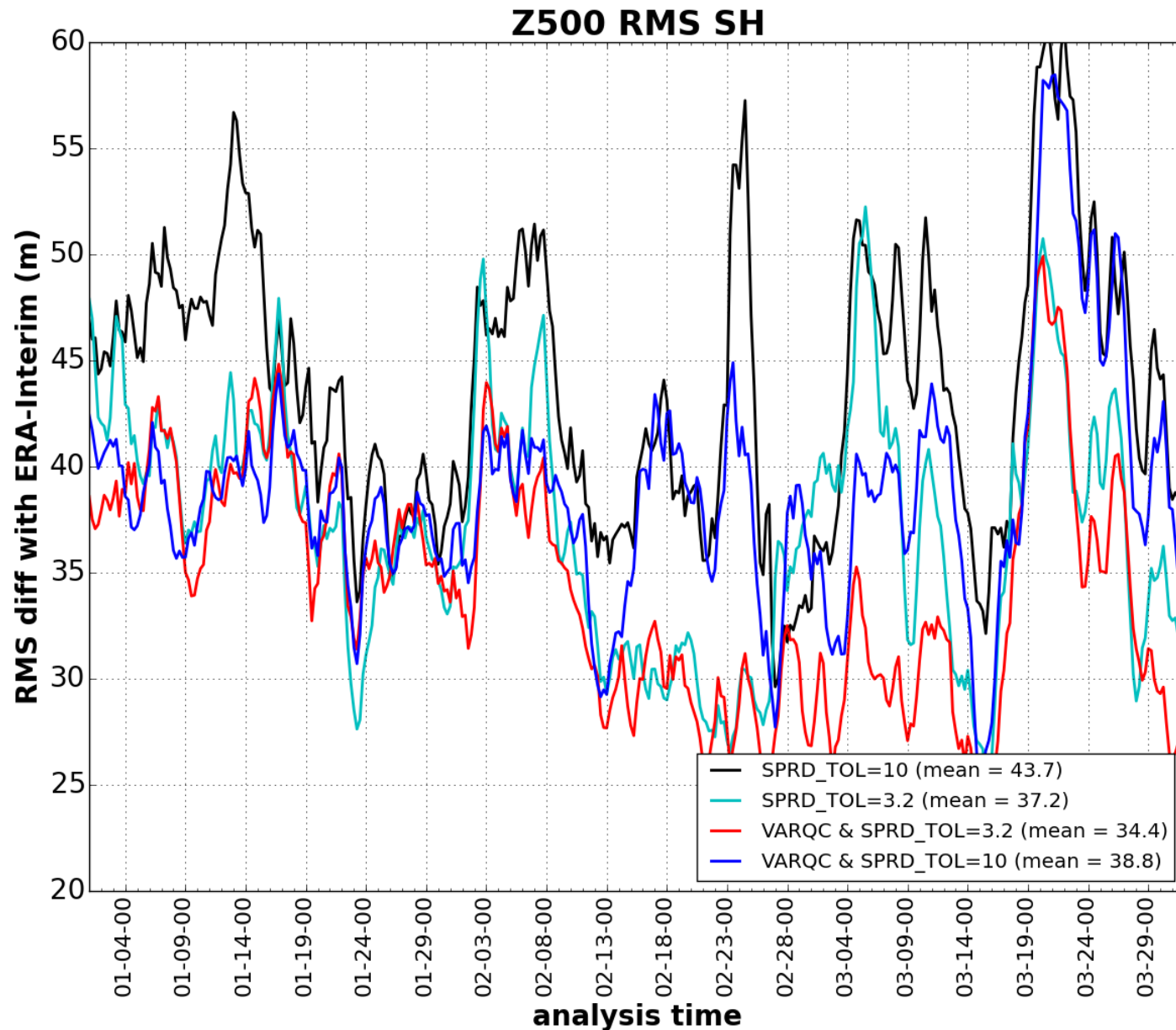Ob rejected of departure greater than **3.2** times sqrt(ens spread + ob error var)

Ob rejected of departure greater than **16** times sqrt(ens spread + ob error var)

Effect of non-gaussian QC in NH

Effect of non-gaussian QC in SH

Varying Localization length scales based upon estimate observation impact: Hypothesis

- The ensemble variance in observation space can be used to estimate a localization scale for each observation.
  - Easily done with a serial EnSRF assimilation algorithm.
  - Specifically, the reduction of ensemble variance in observation space ($\rho = \mathbf{HP^aH^T}/\mathbf{HP^bH^T}=\mathbf{R}/(\mathbf{HP^bH^T+R})$) is inversely proportional to the optimal localization scale.
  - Small ρ -> large variance reduction when ob assimilated -> large signal/noise in ensemble covariance estimate -> larger localization length scale.
- Empirically, we choose $L = L_0(1 - \exp(-(1-\rho)/r))$, where L is the localization scale, L_0 is the maximum allowed localization scale and r is a parameter.
  - Obs are assimilated in order of increasing ρ, ρ recomputed for all obs after each ob is assimilated.

Example: Using a fixed localization scale of 4000 km(obs assimilated in order of increasing ρ), here is the surface pressure increment for the ob with the smallest ρ (the 1st ob assimilated).
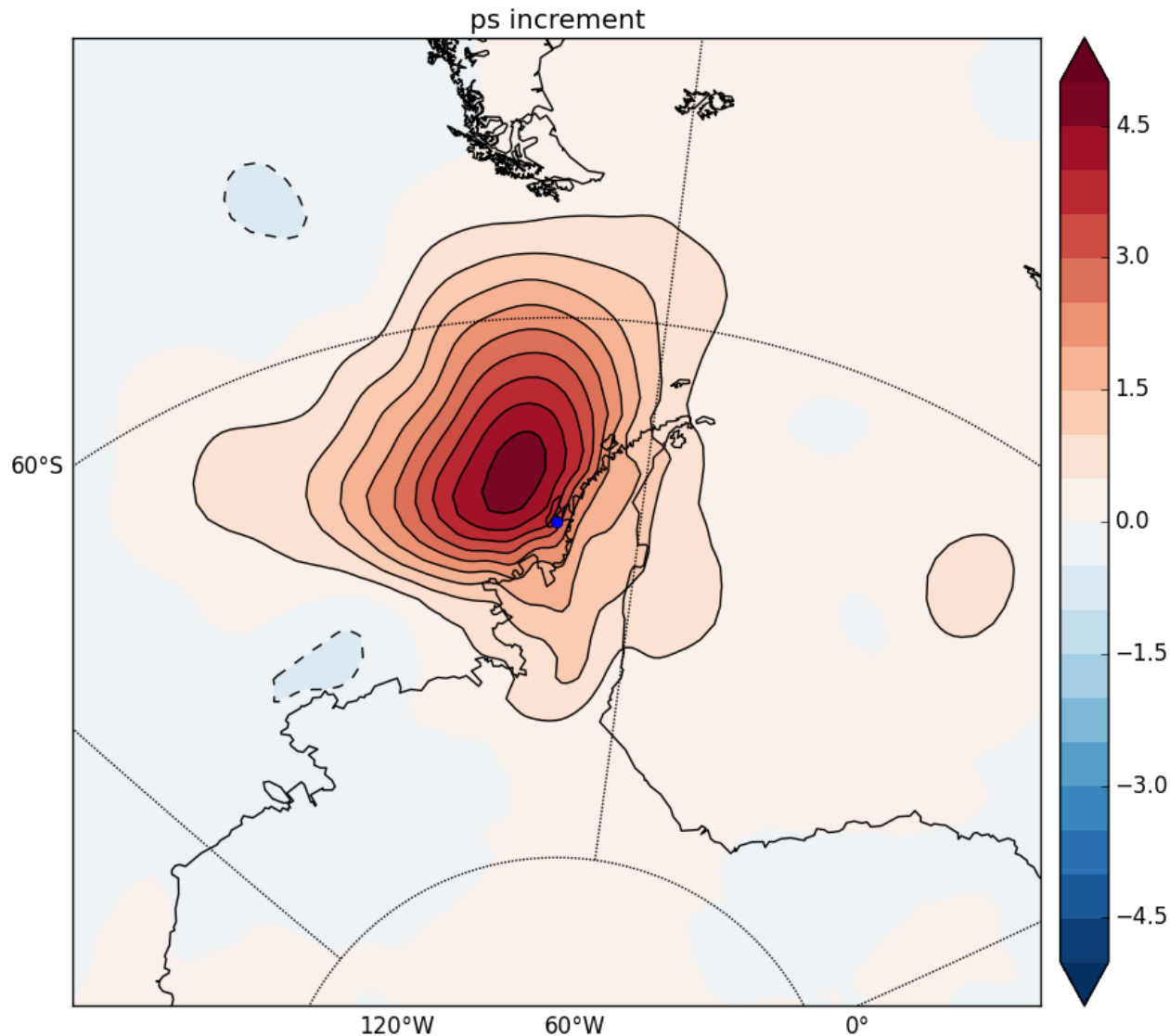

ps increment

Example: Using a fixed localization scale of 4000 km (obs assimilated in order of increasing ρ), here is the surface pressure increment for the 10,000th ob assimilated.

ps increment
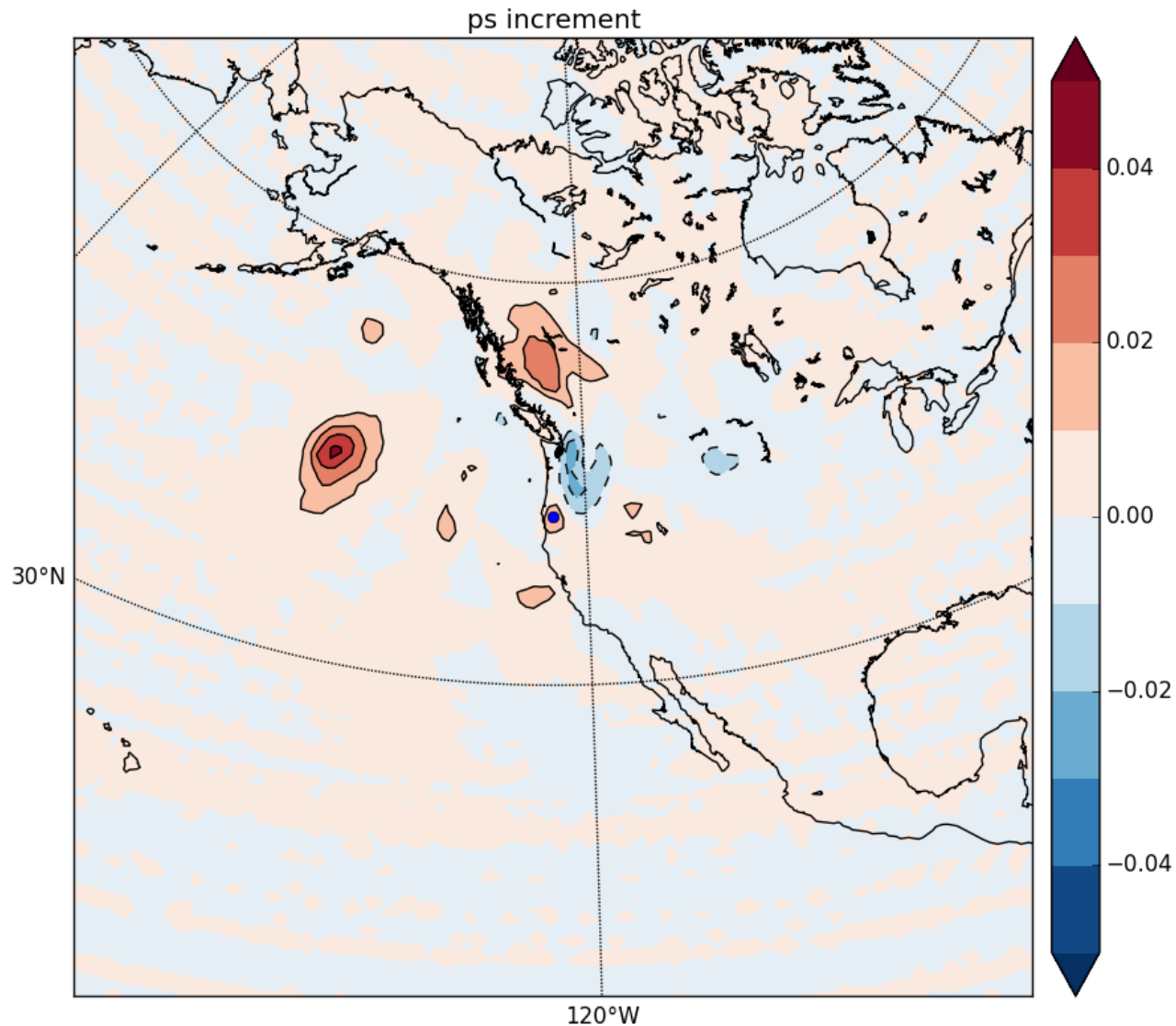
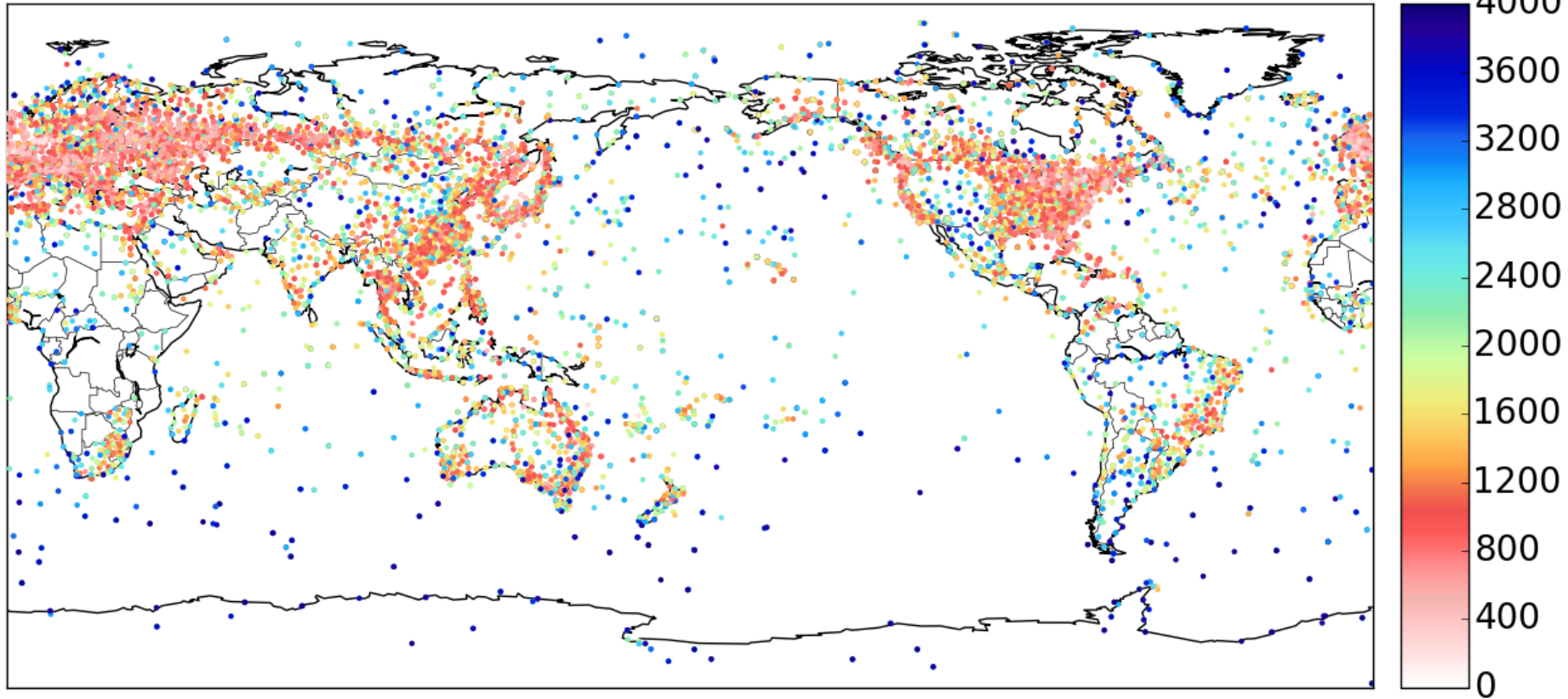# L as a function of ρ , when r=0.2 and L_0=4000km
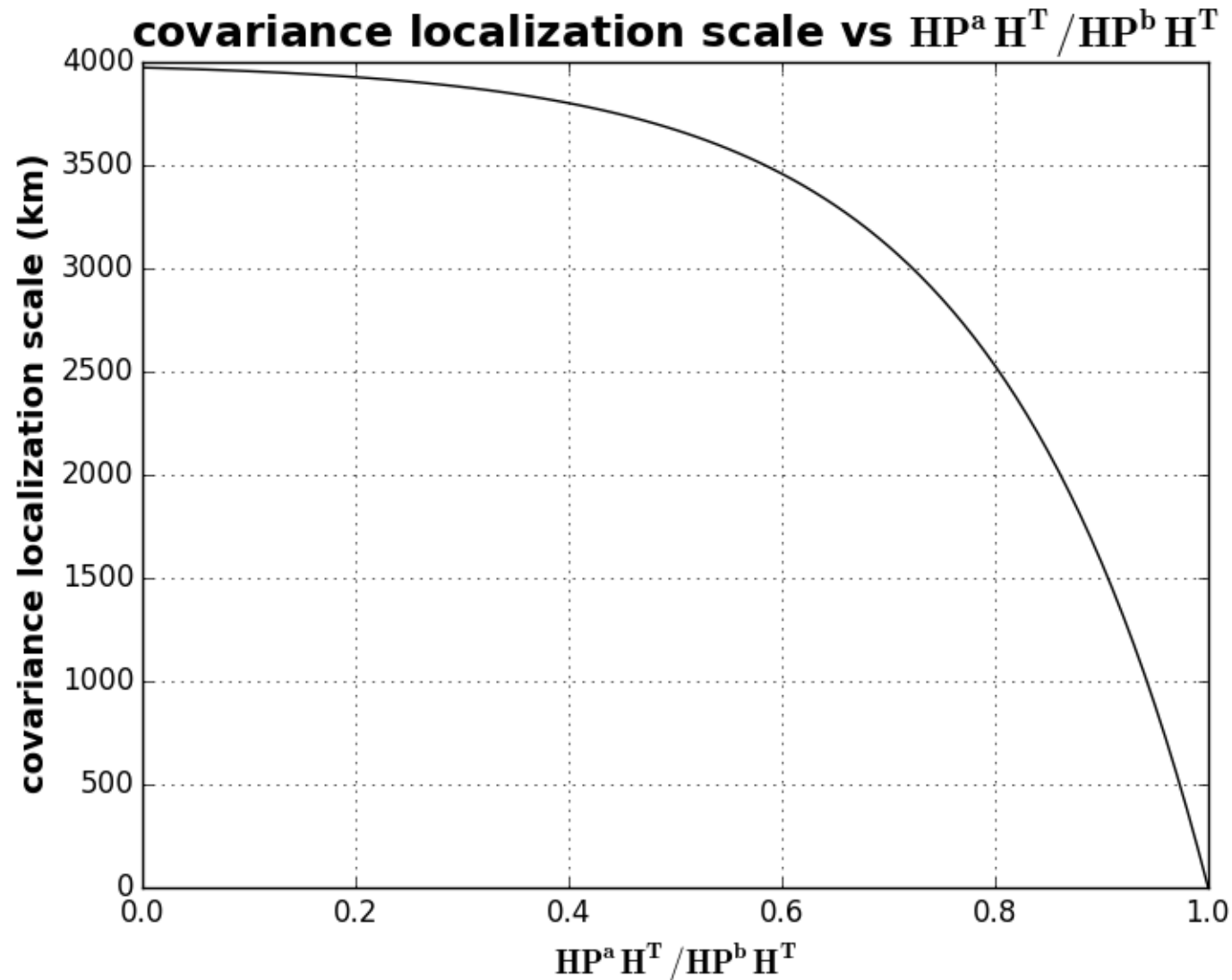
L (red) and ρ (blue) as a function of observation number in the serial algorithm for one assimilation time

$HP^a H^T / HP^b H^T$ (blue) and localization scale (red)
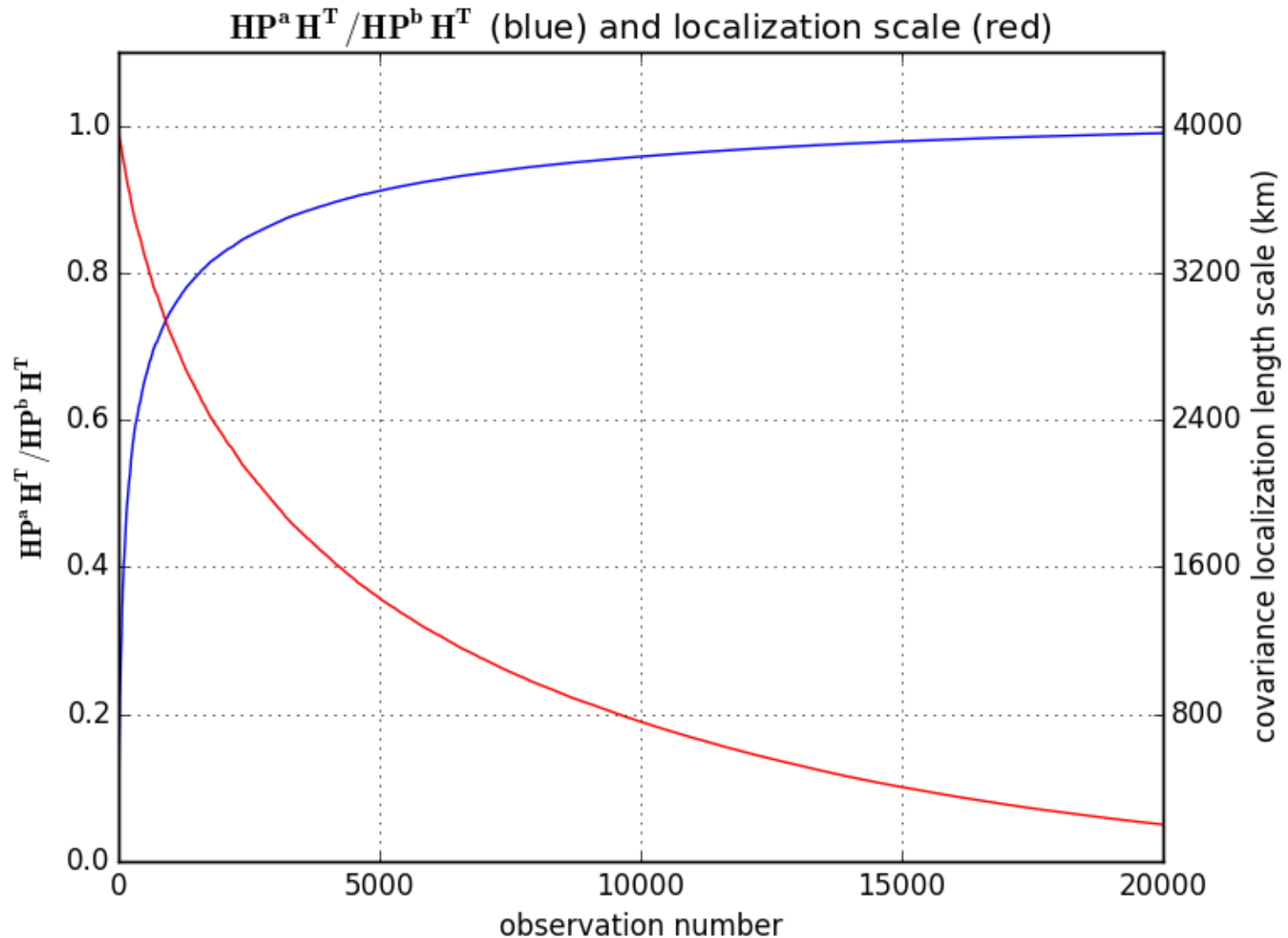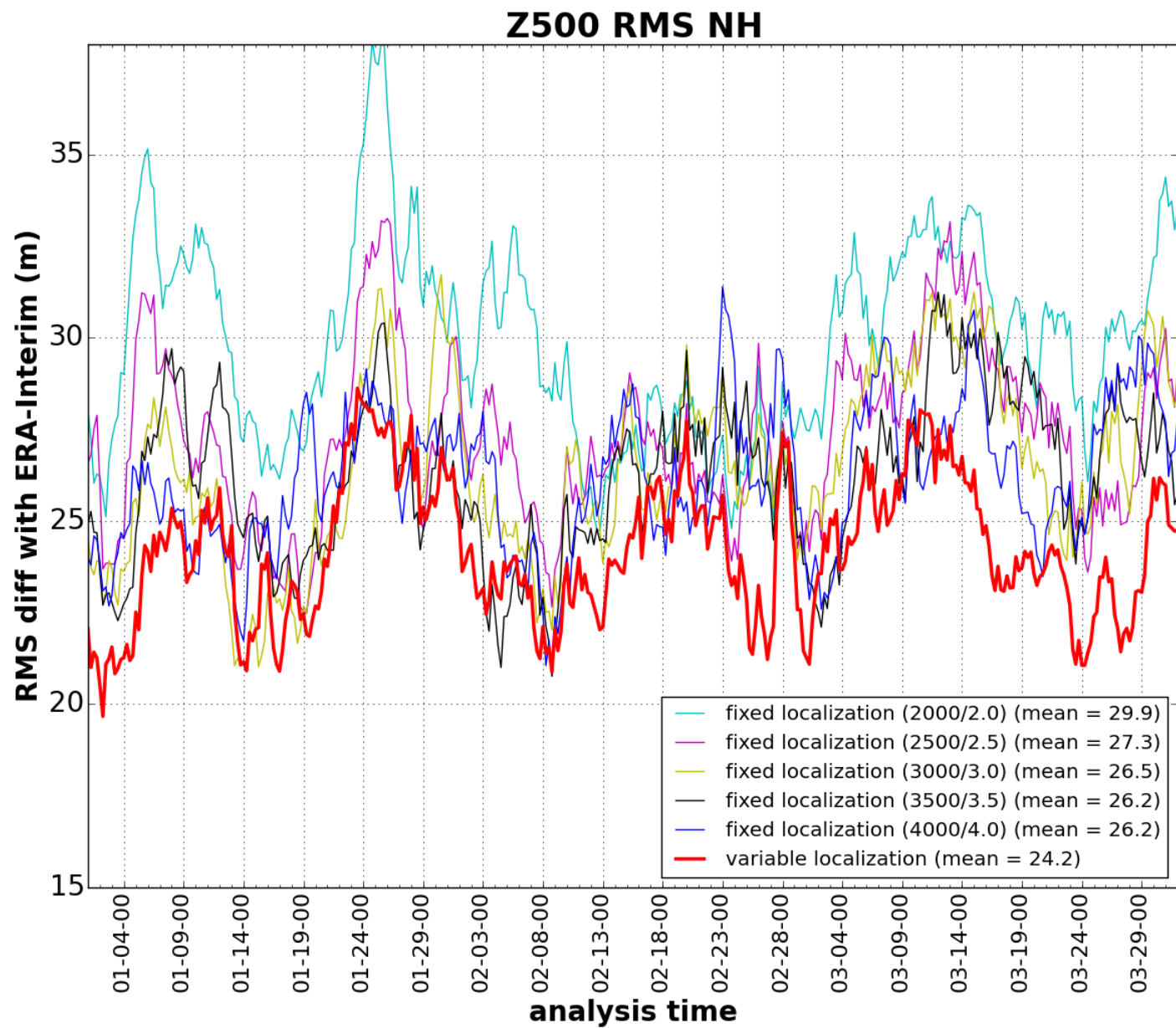
Calculated L for each observation

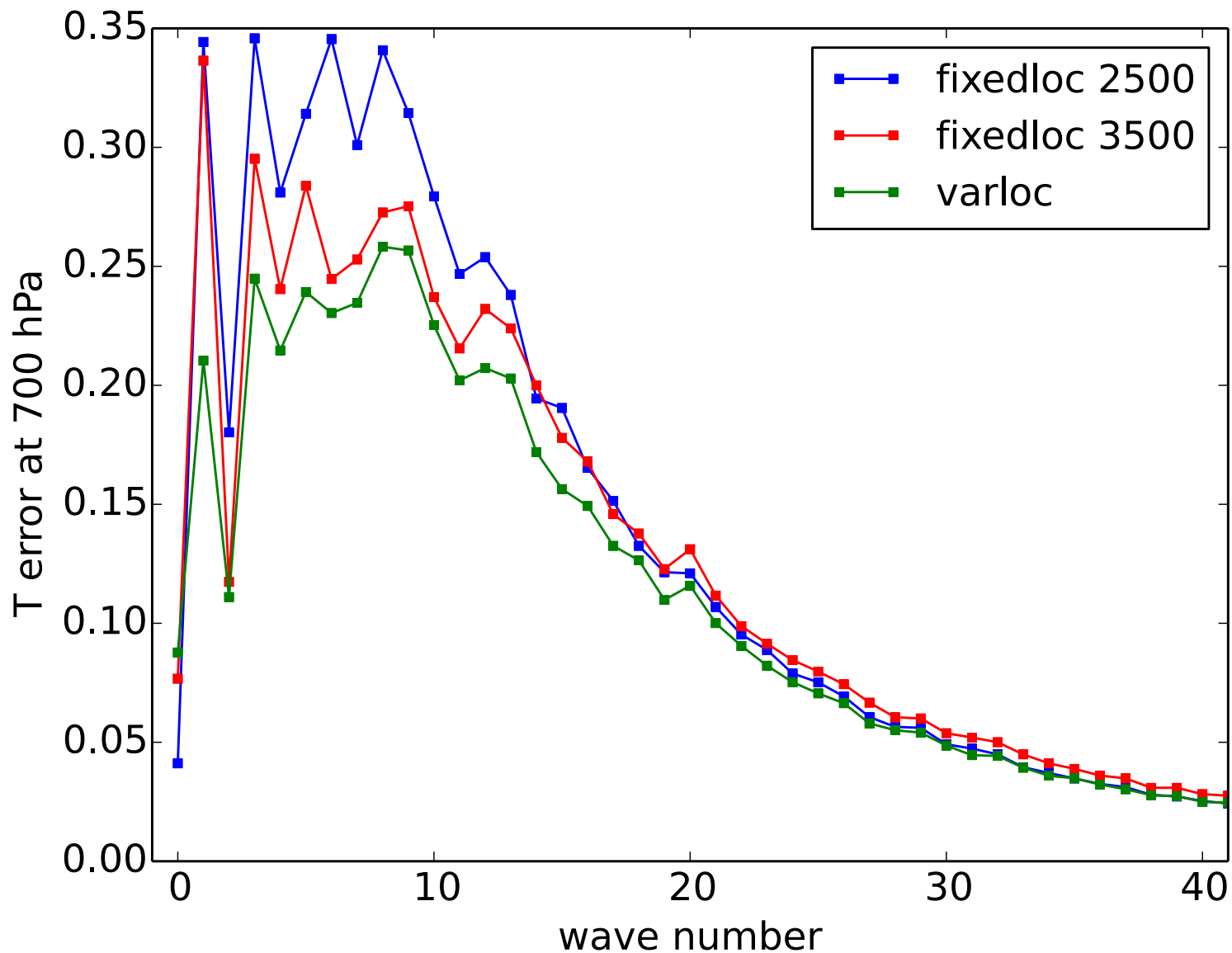covariance localization scale (km) for 2004010100

Experiments

Starting from 1999122000, assimilation run with a 64-member T254L64 GFS ensemble, using only surface pressure observations.

- Observation bias correction and ensemble based quality control used.
- Relaxation to prior spread inflation used with a coefficient of 0.9.
- Fixed localization length scales with the following horizontal (km) and vertical (scale heights) values = 2000/2, 2500/2.5, 3000/3, 3500/3.5, 4000/4.
- Varying localization using r=0.2 and L_0=4000/4.
- Expts run to 2000040100, statistics computed from 2000010100 onward.
- Errors measured against ERA-interim reanalysis.

Z500 RMS NH

RMS diff with ERA-Interim (m) vs *analysis time*

Legend:
- fixed localization (2000/2.0) (mean = 29.9)
- fixed localization (2500/2.5) (mean = 27.3)
- fixed localization (3000/3.0) (mean = 26.5)
- fixed localization (3500/3.5) (mean = 26.2)
- fixed localization (4000/4.0) (mean = 26.2)
- variable localization (mean = 24.2)

Error spectrum at 700hPa T

Varying Localization: Conclusions

- The proposed algorithm performs well in this case, will be used the 20CR version 3.
  - It adapts to spatial and temporal changes in observation density, producing more accurate analyses than can be achieved with a fixed global constant localization scale.
- Experiments will more complete observing systems (multiple observation types) have not yet shown similar improvements relative to a fixed localization scale.
  - Need to normalize observation space variance reduction (ρ) for each observation type to put observations on a more equal footing?
  - For example: perhaps a value of $\rho=0.5$ for an AMSU radiance should not be treated equally with a value of $\rho=0.5$ for a surface pressure observation?

Results from 20CRv3 test for year 2000

RMS 500 hPa Geopotential Height (Northern Hem)
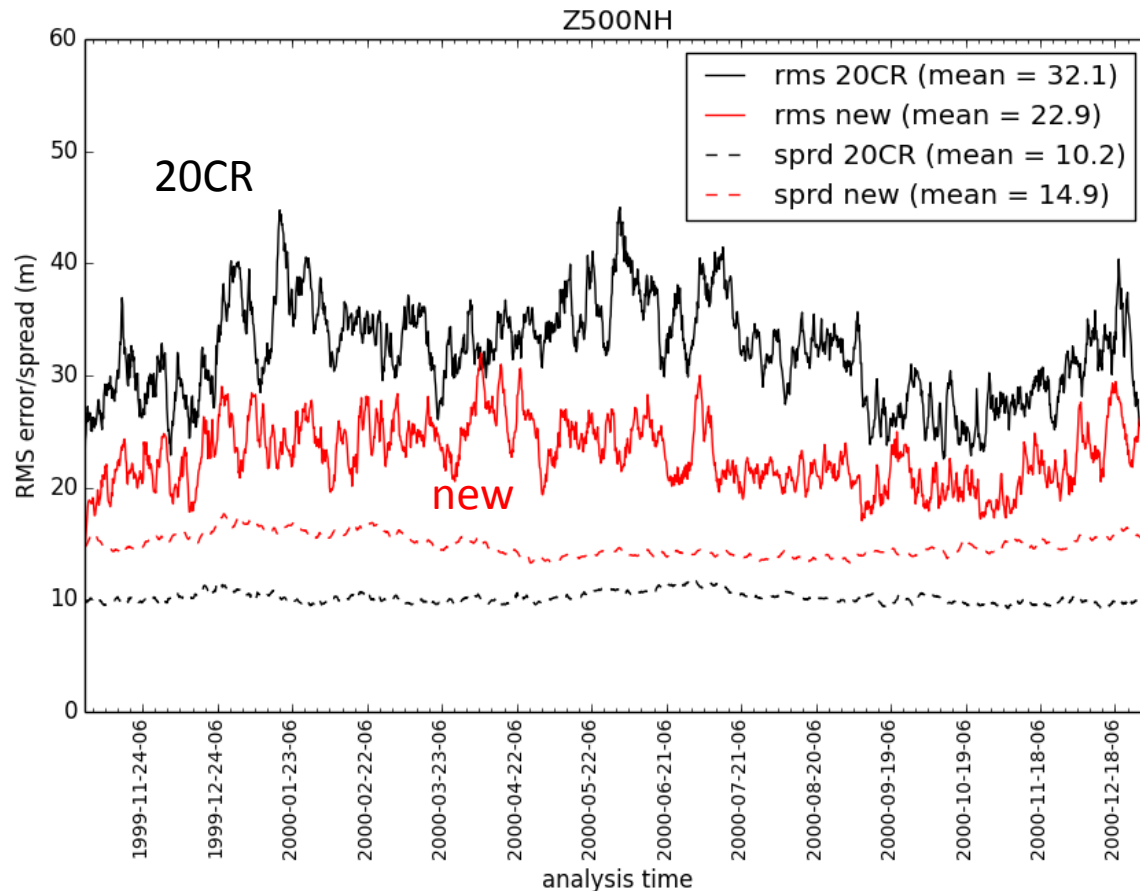
28% reduction in RMS,
spread increased by 45% (now much closer to error)

Summary

- 20CR p_s based analysis system has proven to be a useful testbed for new ideas.
    - Not many obs, so QC is especially important.
    - Covariances matter a lot (easy to see impact).
    - Only a single observation type to worry about.
    - Easy to verify (compare against operational analyses with full observing system).
- Non-gaussian QC and varying localization length scales are novel aspects of the 20CRv3 serial EnSRF system.
    - These, together with higher res model, should produce analyses that are ~25% more accurate in the Northern Hemisphere (relative to 20CRv2).